Empirical Software Engineering
CSE 8340 — Fall 2002

Prof. Jeff Tian, tian@engr.smu.edu
CSE, SMU, Dallas, TX 75275
(214) 768-2861; Fax: (214) 768-3085
www.engr.smu.edu/~tian/class/8340.02f

Module Ib: ESE Guidelines

• Basic Ideas

• Topic Areas or Generic Steps

• Specific Guidelines
ESE Guideline


- Why a guideline?
  - More ESE research activities
  - Maturing of SE and ESE
  - Practical concerns:
    - reader/students
    - researchers/meta-analyst
    - reviewers/editors
    - journals/conferences etc

- Perfect timing for CSE 8340
ESE Guideline: Basis

- Internal basis for the guideline:
  - Research experience in ESE
  - Researcher experience in ESE
  - from both author/reviewer perspectives
  - Other CS/SE work

- External basis for the guideline:
  - Scientific method
  - Implicit guidelines used for emp. studies in other mature disciplines (most natural science)
  - (Explicit) guidelines for emp. studies in other disciplines (e.g., medical)
  - External experts as co-authors

- Result: Preliminary guidelines.
ESE Guideline: Sources

- Authors as information sources:
  - Diverse background
  - Experience in SE/ESE/statistics/others

- Other important sources:
  - Similar guidelines for medical journals
  - Meta-analysis studies
    (studies of empirical studies and results)
  - Papers about statistical applications:
    - positive (guide, ”what should be done”)
    - negative (”what was wrong/to avoid”)
  - Other ”soft” sciences
  - List of specific references in paper
ESE Guideline: Topic Areas

1. Experimental context

2. Experimental design

3. Conduct experiment and data collection

4. Analysis

5. Presentation of result

6. Interpretation of result
Guideline by Topic Areas

- Guideline organization:
  - by topic areas (TAs, or steps)
  - introduction of general ideas, then
  - specific guidelines in the TA

- Notations and shorthands:
  - TA\text{x}: topic area "x" (numbered)
  - Specific guidelines within each TA
  - TA shorthands:
    - C(ontext), D(esign), D(ata) C(ollection), A(nalysis), P(resentation), I(nterpretation)
  - Example: guideline \#4 in TA3 (data collection) is labeled/numbered DC4.
TA1: Context

- Elements of experimental context:
  - background: industry or new
  - research hypothesis
    (if any ⇒ goal-oriented)
  - related research
  - specifics for the above 3

- Types of empirical studies (ES):
  - observational
  - formal experiments
  - other classifications possible
  - re-visit for result interpretation
TA1: Context

- C1: Clearly specify industrial context
  - entities, attributes, measures that capture contextual information
  - in the context of observational or experimental studies

- C1 in observational studies:
  - industry & s/w dev organization
  - staff skill/experience
  - s/w tools/process used, etc.

- C1 in formal experiments:
  - similar info as above
  - do not over-simplify
    (what people call “toy” problems)
TA1: Context

● C2: Hypothesis (if any)
  ▶ clearly state before study
  ▶ theoretical basis for the hypothesis
  ▶ implications?

● C3: if exploratory research:
  ▶ questions to address, and how
  ▶ prior to data analysis

● Comment: goal-oriented assumption
  (GQM as the overall framework)

● C4: describe related research.
TA2: Design

- Elements of experimental design:
  - population
  - sampling technique and rationale
  - treatment (or intervention)
  - bias and sample size

- D1: Identify the population.

- D2: Define sampling.

- Comment: D1 and D2 to ensure that the study generalizable because it is representative of an interesting population.

- Other guidelines (D3-D11) to ensure statistical validity and reduce bias.
TA2: Design

- Statistical validity
  
  ▶ D4: keep it simple
  ▶ D5: define the experimental unit
  ▶ D6: preparation for formal experiment by pre-experiment and adequate sample size

- Reducing bias
  
  ▶ D3 and D10: define treatment (intervention) and describe how.
  ▶ D7: use appropriate level of blinding
  ▶ D8: vested interest (own work)?
  ▶ D9: careful with control
  ▶ D11: outcome related to goal

- Comment: minimize in/external threats to result validity and interpretation.
TA3: Data Collection

- Conducting experiment: domain specific.

- Data collection: common guidelines.
  - DC1: define all measures fully.
    (what we do in CSE 8314)
  - DC2: properly treat subjective ones
  - DC3: accuracy/completeness of DC
  - DC4: (for surveys etc.)
    response rate & representativeness
  - DC5: drop-outs? (for experiments)
  - DC6: other performance measures also

- Comment: DC guidelines to ensure proper, unbiased data supplied to analysis.
TA4: Analysis

- Guidelines independent of the types of analyses performed.

- Types of analyses:
  - classical vs Bayesian: frequency vs pre-post relations
  - parametric vs non-parametric
  - measurement types
  - non-traditional statistical analyses
  - consult statisticians/other experts

- Comment: data/context sensitive guidelines possible, but beyond the scope of general guidelines
TA4: Analysis

• Specific analysis guidelines:
  ▶ A1: careful with multiple testing
    (“torture/fishing” the same set of data?)
  ▶ A2: consider using blind analysis
    (reduce subjective tendencies)
  ▶ A3: perform sensitivity analysis
  ▶ A4: match data with test
  ▶ A5: verify the results

• ”test” = statistical/hypothesis test here

• Comment: proper, unbiased analysis to ensure meaningful results
TA5: Result Presentation

- Presentation of results:
  ensures that others "get it".

- Presentation guidelines:
  - P1: describe/ref. for stat. procedures
  - P2: statistical package used
  - P3: enough details (sig. level etc.)
  - P4: raw data whenever possible
    (independent verification)
  - P5: appropriate descriptive statistics
    (related to details P3)
  - P6: make appropriate use of graphics
    (interesting common errors listed)

- Comment: ensure the readers understand
  the results and the context.
TA6: Result Interpretation

- Interpretation of results: avoid misinterpretation.

- Interpretation guidelines:
  - I1: describe inferential statistics or predictive models (should this be part of the analysis?)
  - I2: stat. significance ≠ practical importance
  - I3: define the type of study (related context...)
  - I4: specify study limitations

- Comment: results meaningful, generalizable? follow-up possible?